卒業論文　2016年度（平成28年度）

期待和了平均順目の評価による
モンテカルロ木探索の麻雀への適用

慶應義塾大学 環境情報学部

細田 航星

徳田・村井・楠本・中村・高汐・バンミーター・植原・三次・中澤・武田
合同研究プロジェクト

2017年1月

卒業論文　2016年度（平成28年度）

期待和了平均順目の評価による
モンテカルロ木探索の麻雀への適用

論文要旨

近年、2人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームでは人間のトッププレイヤと同等かそれ以上の実力を持つコンピュータプレイヤが提案されている。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。本論文では、探索空間が大きくなることでうまく適用できなかったモンテカルロ法について、期待和了平均順目の数理的な評価を利用することで探索空間を小さくし、適用する手法を提案する。この手法の効果を確かめる実験として、多人数性を排除した1人麻雀における和了率の比較を、人間のプレイヤや先行研究AIと行った。またそれをそのまま4人麻雀で対戦させた場合の比較も行う。解析の結果、本手法で構築したAIは1人麻雀において、平均レベルのプレイヤの和了率を出すことができることがわかった。また、4人麻雀に適用した場合、シャンテン数を下げるように打つAIに比べて高いレートを出すことが可能であることがわかった。

キーワード

麻雀, モンテカルロ法, ゲーム情報学

慶應義塾大学 環境情報学部

細田 航星

Abstract Of Bachelor's Thesis　Academic Year 2016

Summary

In recent years, computer games that are equal to or better than the top players of humans have been proposed for games such as chess, Othello, and shogi, which are two-person zero settled perfect information games ing. On the other hand, Mahjong which is a multiplayer incomplete information game has not yet reached it. This is because it is difficult to apply conventional search methods of game trees. In this paper, we propose a method to apply by restricting mathematical evaluation and applied parts of the Monte Carlo method which could not be applied successfully by increasing the search space. For evaluation, comparison of the rate of winning in one mahjong excluding multiplayer was done with human players and previous studies AI. We also make comparisons when we played it with 4 mahjong as it is. As a result of the analysis, we found that the AI constructed by this method has comparable ability to that of the average level player.

Keywords

Bachelor of Arts in Environmental Information

Keio University

Wataru Hosoda

目 次

図 目 次

表 目 次

第1章　序論

1.1　本論文の背景

近年、2人零和確定完全情報ゲームであるチェス、オセロ、将棋といったゲームではAIの研究がめざましく、人間のトッププレイヤと同等以上の成績を記録している。一方で多人数不完全情報ゲームである麻雀では未だそれに及ばない状態である。これは従来のゲーム木の探索手法を適用するのが難しいためである。多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、先行研究 [10][9] でも多数行われている通り、相手を考慮しない 1人麻雀プレイヤに対して和了率を求める評価を採用する。麻雀を行う時の人間の思考は1 人麻雀をもとにしているため、1人麻雀から4人麻雀への拡張は容易だと考えられる。

1.2　本論文が着目する課題

1人麻雀においてある局面からゲーム木を展開しようとした場合、次にどの種類牌をツモるかはわからないため、ランダムに決定しノードを展開していくことになる。これを繰り返し行うと探索空間が大きくなりすぎるため、手の選択が難しくなる。これを対処するために、あらゆる手法が提案されている。このような場合に木を展開せずに有望な手を選択する手法として、Upper Confidence Bound (UCB) [1] がある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。これらは麻雀について適用した例が報告されている [10] が、いずれの方法も平均プレイヤーに満たない成績となった。

本研究では、モンテカルロ法の問題点である、麻雀に適用すると探索空間が大きくなりすぎてうまく適用出来ない問題に対して、探索空間を小さくするために、2つの解決策を行う。一つはモンテカルロ法のシミュレーションを直接牌姿に当てはめるのではなく、牌姿ごとの局面で数理的に比較できる内容を数理計算を用い削減する方法である。もう一つは、三向聴以下に対してはシャンテン数を下げるように打ち、モンテカルロ法を用いる部分を限定する方法である。麻雀の牌理に関しては、上がりに近づく選択ほど重要である [11] ため、聴牌に近い打牌選択においてモンテカルロ法を適用することで、成績に影響を大きく与える部分でモンテカルロ法の効果が発揮できることが期待される。

1.3　本論文の目的

本研究では和了率を理論値計算によって評価することと和了に近い段階へのモンテカルロ法の部分適用によって、麻雀に対してモンテカルロ法をそのままうまく適用できない問題に対して解決策を提案する。多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、先行研究 [1,2] でも多数行われている通り、相手を考慮しない 1 人麻雀プレイヤに対して和了率を求める評価を採用する。従来のモンテカルロ法を麻雀に適用した例では平均プレイヤーレベルの実力が出ていないため、本手法によってそれらを上回ることが期待される。

1.4　本論文の構成

本論文の構成について述べる。2 章では、不完全情報ゲームの関連研究について述べる。3 章では、1 人麻雀にモンテカルロ法を適用する際に、適用範囲を限定し、比較内容を数理的和了率によって評価する手法を提案する。4 章では、3 章で提案した手法で実装した AI に 1 人麻雀を打たせ、各種パラメータを比較する。また、4 人麻雀で打った際の成績の評価も行う。5 章では、本手法によって得られた知見について述べる。

第2章　背景と課題

本章では不完全情報ゲームの関連研究を述べ、麻雀の関連研究がどのように行われているかを述べる。

2.1　ゲーム木探索の背景

探索の方法として一般的なのは、「mini-max 探索」と呼ばれているもので、自らの手はできるだけ高い評価に、そこから想定される相手の候補手はできるだけ低い評価になるように探索していく。その結果を理想的な順番に並べ、可能性の低い選択肢を切り捨てる「$\alpha\beta$ 枝刈り」を行うことで、探索を効率化させることもできる。

ただし不完全情報ゲームにおいてはこの方法では探索空間が広がりすぎて精度を保つことが出来ないので、別の解決策を考える必要があった。

2.2　不完全多人数情報ゲームの例

プレイ人数に着目した研究としてはポーカーを用いたもの [4],[5] がある。これらの手法では 2 プレイヤでのゲームを強くしてから多人数に適応する方法がとられている。ポーカーは 2 プレイヤであればナッシュ均衡戦略を用いることで世界チャンピオンに勝っているが、ポーカーは 2 人から 10 人程度で参加可能なゲームであり人数が増えると状態数が指数関数的に増大するためナッシュ均衡戦略の計算は難しい。そこで 2 人で行われたナッシュ均衡戦略を 3 人限定で拡張する方法、プレイヤの行動を削減、抽象化することでより少ない人数の少ないゲームを想定する方法がとられた。

2.3　麻雀の例

麻雀においての AI 研究については機械学習を用いた水木ら [6] や北川ら [7] の研究が報告されているが、いずれも平均プレイヤの実力に至っていない。一方、水上ら [?] は 1 人麻雀プレイ ヤの学習に牌譜との一致を目指した平均化パーセプトロンを用い、平均プレイヤ以上の実績を収めている。

水上らの研究より、麻雀においてもプレイヤの人数を削減してから多人数に適用する方法は有効である可能性が高く、以後ゲームを単純化した 1 人麻雀に対しての研究が報告されている。

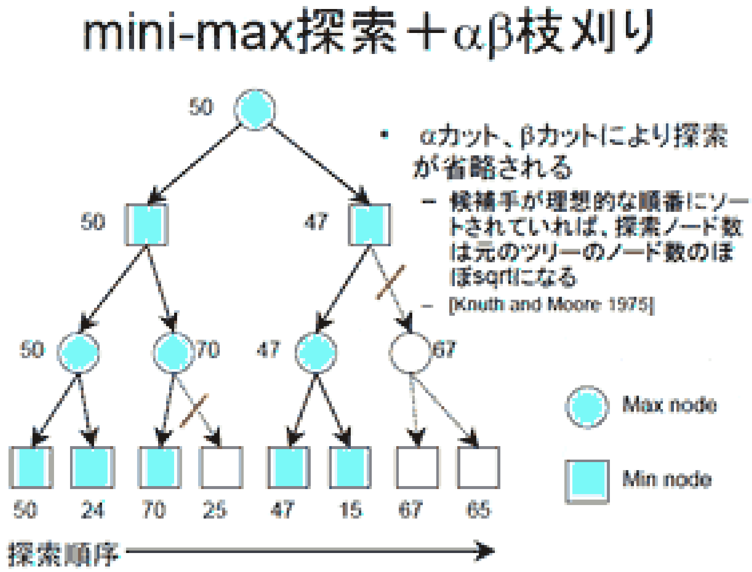

図 2.1: mini-max 探索

2.4 1 人麻雀について

多人数ゲームでは状況によってプレイヤの行動の目的が違うといった難しさがある。この難しさを解消するために、相手を考慮しない 1 人麻雀プレイヤを考える。この 1 人麻雀プレイヤのルールは、相手プレイヤを考えないため、相手プレイヤーに点数を支払う事による放銃や被ツモ失点などを考えない。また、相手プレイヤーによる捨て牌が存在しないので、鳴きや栄和を考えない。リーチについても、和了率の観点では不要なため、考えない。門前でツモを繰り返し行うだけのシンプルなシステムである。先行研究 [?] 評価を合わせるため、ツモの回数は 27 回とした。

表 2.1: 本研究の 1 人麻雀プレイヤのルール

アクション	4 人麻雀	1 人麻雀
和了	ツモ和了、ロン和了どちらも可	ツモ和了のみ
点数	失点や和了の点数を考慮する	点数は考えない
リーチ	可能	なし
鳴き	チー、ポン、カンが可能	なし
終局	誰かが和了するか、ツモ山 70 枚がなくなるまで（136-14-13*4)	和了するか、27 回ツモるか

また 1 人麻雀の成績の評価においてはその点数の推移よりも、和了率自体が重要だと考えられている。したがって、点数という複雑なパラメータを省いた和了率に注目したルールとなっている。その理由としては、以下のような理由が述べられる。

まず、4 人麻雀では平均順位の低いプレイヤほ ど、平均和了点低く、和了率が高いことが統計で明ら かになっている [8] したがって強いプレイヤは点数よりも和了率を重視して打っているこ

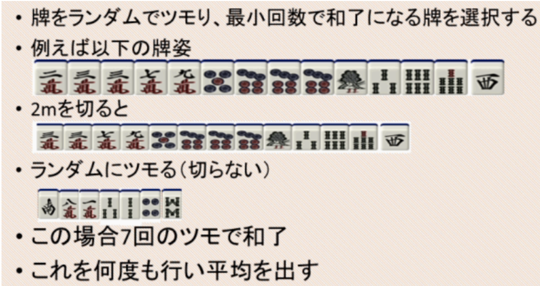

図 2.2: 麻雀を打つ機械の牌効率アルゴリズム

とがわかる。

上手いプレイヤの平均和了点が低くなる理由として手役を無理に狙った打ち方をしないためである。また麻雀の点数の特性上、満貫までの点数は指数関数的にその得点が増えていくが、満貫以上の場合は線形に近くなる。したがって、難易度に対して望める点数が割に合わない高い和了を狙うより、比較的点数の低い点数で多く和了することが効率が良い。また、失点の観点からも、４人麻雀では自分が放銃をしない場合も他プレイヤーの和了によって被ツモ失点を被ることがある。これに対しても、和了すれば他プレイヤーが和了することができないため、相手のチャンスをつぶすといった意味でも和了率の高さは大事である。

2.5 麻雀についてモンテカルロ法を応用した研究

2.5.1 モンテカルロ法

水上ら [?] の研究の前衛である、「麻雀を打つ機械」[3] の牌効率のアルゴリズムでは、モンテカルロ木探索を用いている。そのアルゴリズムの概要を図??に示す。ツモを入れた状態で１４枚の牌姿から、それぞれ切ることのできる牌を切った場合を一つのノードとする。この場合深さ１の１４つのノードができることになる。この後、それぞれのノードの状態から、ツモることの可能な牌をランダムにツモっていくシミュレーションを行う。この場合ツモる事のできる牌とは、場や自分の手牌に情報として見えている牌をゲーム開始時に存在する麻雀牌セットの中から抜いたものである。通常麻雀のルールでは、牌をツモることで次に捨てるという動作が発生するが、ここではそれを行わない。これをしないことで、探索空間が膨大に膨れ上がるの防ぐ効果があり、特定のシミュレーション時に最速で上がるための手順を特定することが可能となる。ツモを連続で行っていき、それまで手牌に残っていた牌の組み合わせで和了することができる場合となるまでシミュレーションを行い、上がった場合にプレイアウトとなる。このシミュレーションを各ノードでそれぞれ十分な回数行い、プレイアウト発生するまでのツモった回数の平均とり評価値とする。最後に、各ノードの評価値を比較し、最もこの値が少なかったノードが最適な打牌と考え選択することになる。プレイアウトまでにツモった回数の平均が少なければ少ないほど、より早く和了できることが期待されるからである。

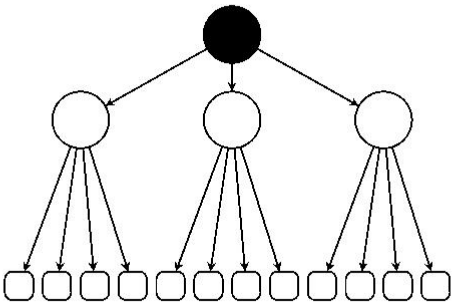

図 2.3: モンテカルロ木探索

2.5.2 UCB1

以上で述べたモンテカルロ法の問題点は、各ノードに対して行うシミュレーションがそれぞれ膨大なため、プレイアウトまでの回数が多く精度が落ちてしまうことである。この問題を解決するために、提案されている手法の中に、Upper Confidence Bound (UCB) [1] を用いたものがある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。有望な手に対して多くのプレイアウトを実行することで、先に述べた通常のモンテカルロ木探索よりも精度を上げることができる。以下、このアルゴリズムを詳しく説明する。

通常のモンテカルロ木探索では、ノードごとに期待される手の良さを途中で判定していないため、全てのノードに対して同じ回数のプレイアウトを行うことになる。したがって、良い手が期待できないノードに対しても多くのプレイアウトを割り当てることになり、無駄なシミュレーションを行ってしまう事が考えられる。また、ノードによっては手の良さを決定する分岐がプレイアウトに対して大量にある場合、他のノードと同じ回数のプレイアウトでは十分な評価が出来ない可能性もある。このような問題に対して、UCB では、UCB 1 値を用いることによって各ノードの有望さをその都度確認し、全体のプレイアウトの回数や特定のノードで行われたプレイアウトの回数を考慮して、プレイアウトをどのノードに数多く割り当てるかということを判別することを考える。UCB 1 値は、x_j は子ノード j の平均報酬, α は定数,n は親ノードの探索回数, n_j は子ノード j とした時、

$$UCB1 = \overline{x_j} + \alpha\sqrt{\frac{2\log n}{n_j}} \tag{2.1}$$

で表される。式 2.1 の右辺の第 1 項は平均報酬を, 第 2 項は信頼度を示している。従来のモンテカルロ木探索では第 1 項の平均報酬をそのまま全てのプレイアウトが終了するまで行った値として利用していたが、UCB1 ではここで第 2 項の信頼度を用いることで全てのプレイアウトが終了するよりも前に信頼できる範囲を推定する。そうすることで、多くのプレイアウトを最後まで割り当てなくとも、一定の信頼度で少ないプレイアウトでそのノードが有望かどうかを判別することが可能になる。信頼度はそのノードのプレイアウト回数が少ないと大きく, 多いと小さくなる。UCB1 値を用いることで UCB1 ではよ り有望そうな手に対して多くのシミュレーションを行う事 ができる。

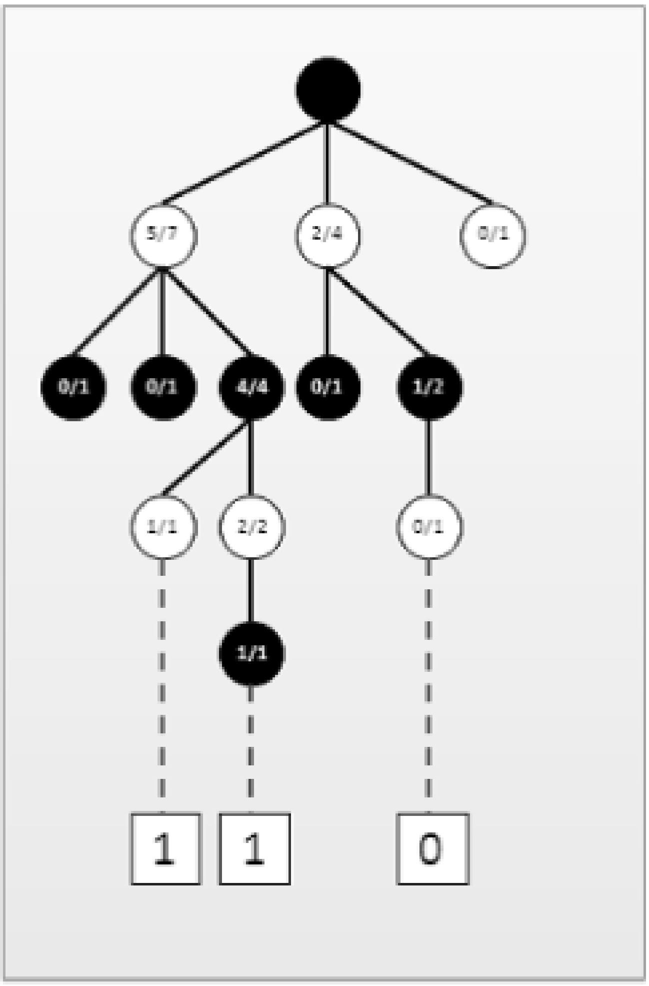

図 2.4: UCB を用いたモンテカルロ木探索

2.5.3 Lin UCB

UCB1 は各局面ごとにこの手順を実行するが, 各局面を完全に異なる局面として扱うため, 他の局面の探 索において得られた情報を共有することができないという 欠点を持つ. この欠点を補うアルゴリズムとして期待され ているのが LinUCB である.

UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。

LinUCB [5] は, UCB を局面を特徴で表すことができる ように拡張したものであり, 牌譜の局面からの教師あり学 習や異なる探索の結果の共有ができる手法である。LinUCB はプレイアウトを行う子ノードを選択する評価値の計算に, 重みベクトル, 特徴ベクトル, 特徴の頻度を表す相関行列を 用いる. 計算により求めた評価値が最大となる子ノードに 対してプレイアウトを行い, 共通で保持する重みベクトル を更新する. 重みベクトルは, 選択したノードの特徴ベクト ルの各項にプレイアウト結果の報酬を乗じた値を, 重みベ クトルの対応する項に足し込んで更新する. この更新によ り, 高い報酬を得た特徴は大きな重みを, 低い報酬の特徴は 小さな重みを持つようになるため, 当該ノードだけでなく, 同様の特徴を持つ他のノードの評価値も更新される. これ により異なる局面で得られた情報を共有し, 利用すること ができる. また LinUCB は探索を行った結果を重みベクト ルとして記録しておくことで, 事前学習の結果を用いる探 索として利用することも可能である. LinUCB のアルゴリズムを Algorithm 1 に示す. ただし xt,at はプレイアウト回数が t 回目の時の選択肢 a の特徴ベ クトル, A は特徴の共起を含めた頻度を表す相関行列, b は各ノードごとの報酬の総和を表すベクトル, θ t は重みベク トル, pt,a は選択肢 a の評価値, α はバランスパラメータ, rt は報酬を表している. rt はプレイアウトにより与えるか, 学習データにより与えるものとする. LinUCB は 4 行目で 前回のプレイアウト結果を反映して重みベクトル θ t を更 新する. 7 行目で重みベクトル, 特徴ベクトル, 相関行列を 用いて各ノードの評価値を計算し, 9 行目で評価値が最大 となるノードを選択する. 10 行目でプレイアウトを行い報 酬を受け取り, 11 行目, 12 行目で A と b の更新を行う. LinUCB は Algorithm 1 の 7 行目に示したように, 式 (2) により各ノードの評価値を求める. 式 (2) の第 1 項は各ノー ドの平均報酬を計算しており, 第 2 項で信頼度を計算して いる.

$$pt, a = \ \theta \ Ttxt, a + \ \alpha \tag{2.2}$$

また, UCB で用いている評価値も, 平均報酬と信頼度の和 で計算される. つまり LinUCB と UCB は本質的に等しい 計算をしていることが分かる. また, LinUCB の第 2 項は データ数の増加により十分速く小さくなることが保証され ている [5]. 以上より, LinUCB は UCB と近い運用を行うことができると考えられる.

これらは麻雀について適用した例が報告されている [10] が、いずれの方法も平均プレイヤーに満たない成績となった。

2.6 本論文が着目する課題

1 人麻雀においてある局面からゲーム木を展開しようとした場合、次にどの種類牌をツモるかはわからないため、ランダムに決定しノードを展開していくことになる。これを繰り返し行うと探索空間が大きくなりすぎるため、手の選択が難しくなる。これを対処するために、あらゆる手法が提案されている。このような場合に木を展開せずに有望な手を選択する手法として、Upper Confidence Bound (UCB) [1] がある。UCB ではある局面から考えられる全ての手に対して, よい結果 を返しそうな手を重視しつつ、何度もプレイアウトと呼ば れるランダムシミュレーションを行うことで 最善の手を決定する手法である。UCB は局面ごとにこの探索を実行するが, 各局面を完全 に別の局面として扱うため他の局面の探索において得た情 報を利用することができない。この問題を解決するために Linear UCB (LinUCB) [2] という手法が提案されている。これは、局面を特徴で表す ことで、対象とする局面が異なっても, それまで対象とした局面の情報を利用できるといったものである。以上のような対策が考えられてはいるが、 1 人麻雀について適用した結果平均プレイヤーの実力までに至っていない。囲碁においては一定の成果を上げたものの、麻雀ではうまく適用できない問題に対しては、麻雀の場合はより探索の空間が大きく、UCB 1 値や LinUCB を指標とするノードの展開の方法ではうまく適用できていない可能性がある。したがって本研究では、ノードの展開を行う際に指標とするものを麻雀というゲームの性質上の状況のパラメーターを用意し、それを使うこととする。

第3章　提案手法

3.1　探索空間を限定する手法の提案

3.1.1　和了率の評価によるノードの展開

本研究では、モンテカルロ法の探索空間が大きくなりすぎる問題に対して、特定のノードの牌姿に対する期待和了順目を次の節で説明する数理モデルを使って有望だと思われる手を探索するようにする。1人麻雀において特定の牌姿の期待和了順目を途中で評価できることは、より有望な手を探索できる可能性がつながるため、UCB 1などの評価値よりも手の有望さを評価する精度を上げることができると期待される。

3.1.2　1人麻雀プレイヤにおける和了率の数理的評価

国士氏の研究により、ある手牌における聴牌までの平均消費順目は、それぞれの変化での平均消費巡目のそれぞれの変化する確率での単純な平均で与えられることがわかっている。[3] すなわち、元の成功率がp、手変わりする確率がq,r、手変わり後の成功率がQ,Rならば、

その手の向聴が進むまでの平均消費巡目は

$$\frac{p\frac{1}{p} + q\frac{1}{Q} + r\frac{1}{R}}{p+q+r} \tag{3.1}$$

と表される。

本論文ではこれを和了時までの式に拡張する。まず、聴牌までの平均順目と聴牌後の和了までの平均順目の合計がすなわち全体の平均消費順目であることを証明する。次に、それを合成した結果の数式がどうなるかを書く。

考察の結果このようになる。

$$\frac{p\frac{1}{p} + q\frac{1}{Q} + r\frac{1}{R}}{p+q+r} \quad + \text{和了期待順目} \tag{3.2}$$

また、(現在の平均テンパイ巡目)=(次巡の平均テンパイ巡目)+1　であることから、各ノードが手替わり率とテンパイ率を値として持つ深さ1の木構造で記述できる。したがって、同様に、n

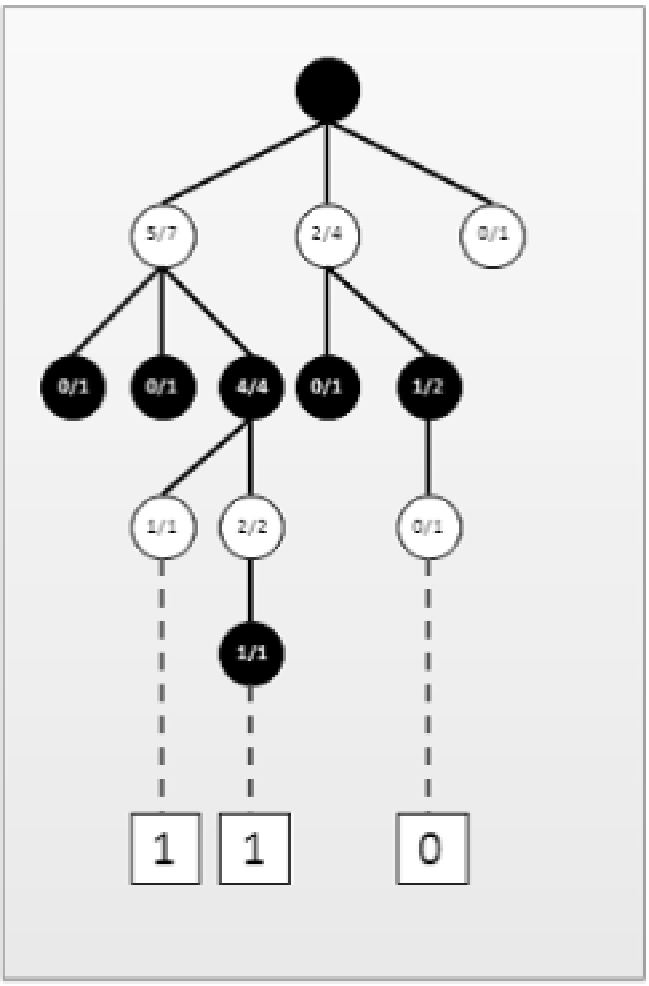

図 3.1: 和了率評価を用いたモンテカルロ木探索

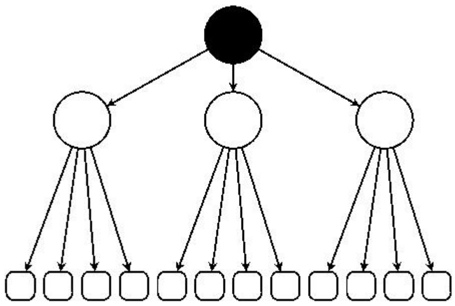

図 3.2: 図の説明

手先の手替わりまで考えると深さ n の木構造で記述することができる。n を大きくすることにより、精度の高い和了率を近似することが可能となる。

図 Q

本研究では、この式 (3) をそれぞれの牌姿に当てはめ、その計算結果が高くなるようなノードを探索する。これにより、従来の手法では何回も同じ牌姿の変化をシミュレートしなければいけない問題があったが、（３）式の評価によってその多くが削減できる。したがって、この方法によって精度が高くなることが期待される。

3.1.3　二向聴以下の牌姿に対してモンテカルロ法を適用する

麻雀において、和了までの手順の中では、シャンテン数が小さい時点での選択が成績に影響を与えやすいことがわかっている [11] したがって、成績向上のためには和了系に近い部分において戦略を改善することが成績に影響を与えやすい事がわかる。

また、モンテカルロ法の問題点は、前述したとおり麻雀に適用すると探索空間が大きくなりすぎることが問題であった。しかしこれを和了系に近い部分に適用することで、探索空間を小さく削減することができるため、適用することができるようになる。本研究の手法である与えられた牌姿における和了率の近似式（３）を適用する際にも、シャンテン数が大きすぎる場合については正確に見積もることが難しいため、このような限定的な部分への適用が重要である。

本論文では二向聴以上の部分にこのモンテカルロ法を適用することで、従来の問題点を解決する。

第4章　評価

4.1　実験概要

4.2　１人麻雀プレイヤによる和了率の比較

提案手法で提示した、期待和了平均順目の評価によるモンテカルロ木探索を 1 人麻雀に適用し、性能を評価した。期待和了平均順目を n 順先の変化まで評価することによって性能が変わる可能性があるので、１順、２順、３順までの変化を考慮するものをそれぞれ分けて実験した。それぞれに対して 100 局のテストデータと 10,000 局のテストデータを与え、あがることができた局数を計測した。テストデータとは、全ての牌をランダムに並べたものを１００セット用意したデータである。このデータから 13 牌を初期牌として 与え, その後牌を引いて切る動作を 27 回行い, その中であ がれたかどうかを確認した。100 局のテストデータで人間 のプレイヤとの比較を行い, 10,000 局のテストデータで各 手法の性能の評価を行った。

表 4.1: １００局のデータセットでの上がった局数

手法	あがった局数（%)
上級者	51
期待平均順目評価	35
平均プレイヤ	34
UCB1	24
独立 LinUCB	22

4.3　４人麻雀における先行研究とのレートの比較

本研究では 1 人麻雀における和了率の向上を目指すため、和了率の数理的評価とモンテカルロ法を適用する部分を限定する手法をとった。佐藤らは、１人麻雀における和了率を有効牌を数え上げて大きくなるようにすることで、和了率の最大化を図り、これを４人麻雀で打たせレートを取った。同じように本研究手法でも４人麻雀で打たせた結果を比較した。

このように、レートは本手法の方が上回る結果となった。ここで保証安定レートについて比較をする。

表 4.2: オンライン麻雀天鳳一般卓における関連研究の成績

	1 位率	2 位率	3 位率	4 位率	平均順位	試合数
全ツ	0.181	0.216	0.252	0.351	2.77	504
オリ実装	0.237	0.240	0.259	0.264	2.54	834

第 5 章　結論

5.1　本研究のまとめ

本研究では、各牌姿における期待和了平均順目の評価をモンテカルロ木探索に利用することによって、モンテカルロ木探索を麻雀に適用することを提案した。評価として、1 人麻雀の和了率を比較した。先行研究では UCB 1 や LinUCB を 1 人麻雀に適用する例が報告されていたが、これらは和了までのシミュレーションを行っているものであり、探索空間が多く精度が落ちるため平均プレイヤには及ばなかった。本研究の提案手法では、和了率が平均プレイヤ達することが確認できた。また、期待平均和了順目をどれほど深くの変化まで計算するかに関しては、深く計算することでより精度が保たれることが確認できた。ただし、5 巡目以降に関しては大きな違いが見られないことがわかった。4 人麻雀の評価については、シャンテン数を下げるように打った先行研究よりもレートが高いということを示すことが出来た。

5.2　本研究の結論

麻雀において従来の探索方法では探索空間が大きくなりすぎて精度が落ちる問題に対して、期待平均和了順目を用いて和了への確率を簡潔に評価することにより、探索空間を小さくする方法が有用であることがわかった。また、期待平均和了順目はおよそ 5 巡目までの変化を考慮することで十分な精度が期待できることがわかった。最後に、4 人麻雀 AI のアルゴリズムとして、モンテカルロ木探索を期待平均和了巡目を指標に行う手法についてはある程度性能を期待できることがわかった。

5.3　今後の課題と展望

今回の検証では、既存のモンテカルロ木探索を行った研究よりも、1 人麻雀の和了率は高くなることを示すことが出来た。しかし、牌譜から教師信号を学習した方法により和了率を上回ることは出来ていない。この原因としては、期待平均和了順目を用いる方法では、ツモを無限思考回数行える場合の平均順目を最小にするように考えていることである。実際の麻雀では、ツモ回数が限られていて、相手の和了によってもその終局までの巡目が異なる。今後の課題としては、期待平均和了順目に対して、残りツモ回数の変数を導入したモデルを構築し、同じようにモンテカルロ木探索を行うことである。このようにすることで、より 4 人麻雀に近い状態での和了率最大化を目指すことができる。

参考文献

[1] Auer, p., cesa-bianchi, n. and fischer, p.: Finite-time analysis of the multiarmed bandit problem, machine learning, vol. 47, no. 2-3, pp. 235256 (2002).

[2] Li, l., chu, w., langford, j. and schapire, r. e.: A contextual-bandit approach to personalized news article recommendation, proceedings of the 19th international conference on world wide web, acm, pp. 661670 (2010).

[3] nmizu, モンテカルロ法を用いた牌効率, コンピュータで鳳凰卓を目指す part4,http://www.nicovideo.jp/watch/sm25236127(2014).

[4] Risk, n. a. and szafron, d.: Using counterfactual re- gret minimization to create competitive multiplayer poker agents, proceedings of the 9th international conference on autonomous agents and multiagent systems: volume 1-volume 1, international foundation for autonomous agents and multiagent systems, pp. 159166 (2010).

[5] 古居敬大:相手の抽象化による多人数ポーカーの戦略の決 定, 修士論文, 東京大学 (2013).

[6] 三木理斗:多人数不完全情報ゲームにおける最適行動決定に関する研究, 修士論文, 東京大学 (2010).

[7] 北川竜平:麻雀の牌譜からの打ち手評価関数の学習,pro- ceedings of 12th game programming workshop (2007).

[8] とつげき東北. おしえて!科学する麻雀, 2009.

[9] 水上 直紀ほか. 降りるべき局面の認識による 1 人麻雀プレイヤの 4 人麻雀への適用. 第 18 回ゲームプログラミングワークショップ（GPW2013）, Vol. 31, No. 3, pp. 1–7, 2013.

[10] 中張 遼太郎ほか. Linucb の 1 人麻雀への適用. ゲームプログラミングワークショップ, 2013.

[11] ネマタ. テンパイの技術. 勝つための現代麻雀技術論, 2014.